Bayesian optimisation applied to protein-binding peptide sequence optimisation

Jérôme Eberhardt, Markus Lill, Torsten Schwede

Biozentrum, SIB Swiss Institute of Bioinformatics, University of Basel, Basel
jerome.eberhardt@unibas.ch

The ability to probe and alter interactions between macromolecules and small molecule entities with high specificity forms one of the foundations of modern life science. The design of small molecules binding to specific targets proceeds through iterative cycles of hypothesis formulation, molecule generation, biological testing, data analysis, leading to an updated structure-activity relationship hypothesis to initiate a new cycle. Traditionally, this process involves scientists and facilities from several disciplines with manual “hand-over” of the results from each step.

Recent developments in chemistry, biology, Machine Learning (ML) and robotics allow for the first time to envision such an automated closed-loop design-make-test (DMT) environment to design bioactive molecules for life-science and clinical applications. Guided by ML algorithms, the platform would combine fully automated modules for molecular synthesis, purification, and activity testing, to enable the iterative variation and optimization of a bioactive molecule.

As a first step toward fully automated small molecules design, we will start with established automated solid phase chemistry for peptidic compounds, which provides a molecule class of broad biomedical relevance and structural variability. This involves, at the beginning, benchmarking existing methods and checking their applicability, in terms of computational time and prediction accuracy, in the context of a closed-loop environment with no human intervention. For the design and optimisation of peptidic sequences a combination of an active learning method, Bayesian Optimisation (BO), with different sequence- and structure-based initialization strategies to generate the initial peptide population. We demonstrate the capacity of this approach using the Major Histocompatibility Complex (MHC) class I receptor system as a benchmark dataset. Starting with a single peptide-lead sequence in the µM IC50 range and efficiently optimising it to approach optimal binding affinity within just 4-5 DMT cycles. We extensively evaluated its performance, varying conditions and parameters. The developed approach can effectively handle various peptide lengths simultaneously, and also in small batches provide a valuable foundation for peptide optimization in closed-loop DMT environments. Our studies underline the potential of our BO method to efficiently navigate vast peptide sequence spaces, significantly advancing the development of new bioactive peptides. The method, Mobius, is publicly available at https://git.scicore.unibas.ch/schwede/mobius.